An adaptive ensemble Kalman filter for soil moisture data assimilation

Rolf H. Reichle,[1,2] Wade T. Crow,[3] and Christian L. Keppenne[2,4]

Received 24 July 2007; revised 29 October 2007; accepted 4 December 2007; published 21 March 2008.

[1] In a 19-year twin experiment for the Red-Arkansas river basin we assimilate synthetic surface soil moisture retrievals into the NASA Catchment land surface model. We demonstrate how poorly specified model and observation error parameters affect the quality of the assimilation products. In particular, soil moisture estimates from data assimilation are sensitive to observation and model error variances and, for very poor input error parameters, may even be worse than model estimates without data assimilation. Estimates of surface heat fluxes and runoff are at best marginally improved through the assimilation of surface soil moisture and tend to have large errors when the assimilation system operates with poor input error parameters. We present a computationally affordable, adaptive assimilation system that continually adjusts model and observation error parameters in response to internal diagnostics. The adaptive filter can identify model and observation error variances and provide generally improved assimilation estimates when compared to the non-adaptive system.

Citation: Reichle, R. H., W. T. Crow, and C. L. Keppenne (2008), An adaptive ensemble Kalman filter for soil moisture data assimilation, *Water Resour. Res.*, *44*, W03423, doi:10.1029/2007WR006357.

1. Introduction

[2] In the past decade, land data assimilation has been a very active field of research. Assimilated observations include satellite retrievals of surface soil moisture, snow water equivalent, snow cover area, and land surface temperature [e.g., *Andreadis and Lettenmaier*, 2005; *Bosilovich et al.*, 2007; *Dong et al.*, 2007; *Drusch*, 2007; *Reichle et al.*, 2007; *Slater and Clark*, 2006; *Van Den Hurk et al.*, 2002]. The basic tenet of data assimilation is to combine the complementary information from measurements and models of the land surface into a superior estimate of the geophysical fields of interest. In doing so, land data assimilation systems interpolate and extrapolate the observations and provide complete estimates at the scales required by the application, both in time and in the spatial dimensions. Land data assimilation systems thereby organize the observational information into physically consistent estimates of the variables of relevance to data users.

[3] The optimal combination of the measurements with the model information rests on the consideration of the respective uncertainties of each. At times and locations for which highly accurate observations are available, the assimilation estimates will be close to these observations. At times and locations that are not observed (or for which quality control rejected the observation), the assimilation estimates will draw close to the model solution, but will nonetheless be subject to the influence of observations in spatial or temporal proximity of the location of interest.

[4] A key problem in the application of data assimilation methods to land surface problems is the fact that the model and the observational uncertainties themselves are poorly known. It is generally difficult to characterize accurately even basic error information such as the magnitude of the model and observation error variances, let alone spatial, temporal, and cross-correlations and higher order moments. This is a serious problem because data assimilation systems operating with poor estimates of input uncertainties may very well produce degraded estimates of land surface conditions when compared with the model estimates or the observations alone [*Reichle and Koster*, 2002; *Crow and Van Loon*, 2006; *Durand and Margulis*, 2008].

[5] Careful estimation of the model and observation error parameters outside of the data assimilation system is one common approach to alleviate the problem and a necessary first step for any data assimilation application. However, it is typically difficult to obtain a sufficient amount of independent information for such off-line calibration. Furthermore, the true error parameters in the cycling assimilation system may well be different from those in the off-line calibration and may be changing with time. It is sometimes possible to calibrate the input error parameters by conducting a suite of assimilation integrations with various sets of input error parameters and then picking the parameter set associated with the best-performing assimilation experiment (see for example Figure 3 by *Reichle et al.* [2002b]). (Note that this strategy is occasionally called "adaptive" filtering.) It is, however, not computationally feasible for large applications.

[6] Adaptive assimilation approaches have been developed that estimate the model and observation error param-

[1]Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, Maryland, USA.
[2]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[3]Hydrology and Remote Sensing Laboratory, USDA/ARS, Beltsville, Maryland, USA.
[4]Science Applications International Corporation, Beltsville, Maryland, USA.

0043-1397/08/2007WR006357$09.00

eters within the cycling assimilation system based on information from internal diagnostics such as the observations-minus-forecast residuals (or innovations) [*Mehra*, 1970; *Gelb*, 1974; *Dee et al.*, 1985; *Moghaddamjoo and Kirlin*, 1993]. Most theoretical developments for adaptive tuning are based on linear systems with Gaussian errors that operate under stationary conditions. *Gelb* [1974; p. 320] summarizes eloquently that "*in practice [the optimal approaches] may not work as well as theory would predict; other more heuristically motivated approaches may be both computationally simpler and more effective.*" Applications of adaptive methods in atmospheric data assimilation typically follow a covariance matching approach [*Dee*, 1995; *Dee and da Silva*, 1999] and have been used in ensemble data assimilation [*Mitchell and Houtekamer*, 2000; *Anderson*, 2007].

[7] In this paper we modify the adaptive approach of *Dee* [1995] for soil moisture data assimilation. In a 19-year twin experiment over the Red-Arkansas river basin we first demonstrate how poor choices of model and observation error parameters affect the quality of the assimilation products and then show how an adaptive strategy can be employed to mitigate the problem. Section 2 gives an overview of the data assimilation approach and the adaptive strategy used here. In section 3 we describe the experiment setup. Section 4 gives results about the sensitivity of the assimilation estimates to input error parameters and the performance of the adaptive filter. A summary and conclusions are given in section 5.

2. Data Assimilation and Adaptive Filtering

[8] In a data assimilation system, the model-generated soil moisture is corrected toward the observational estimate, with the degree of correction determined by the levels of error associated with each. In most assimilation methods, only the covariance information of the model and observation errors is used under the assumption that errors are Gaussian. Sequential data assimilation algorithms known as Kalman filters [*Gelb*, 1974] explicitly propagate error covariance information from one update time to the next, subject to possibly uncertain model dynamics, and produce estimates of the uncertainty in the assimilation products. Adaptive filtering approaches are designed to estimate the input uncertainties along with the geophysical quantity of interest in the cycling assimilation system.

2.1. The Ensemble Kalman Filter

[9] In this paper we use the ensemble Kalman filter (EnKF), a Monte-Carlo variant of the Kalman filter [*Evensen*, 2003]. The idea behind the EnKF is that a small ensemble of model trajectories captures the relevant parts of the error structure. Each member of the ensemble experiences perturbed instances of the observed forcing fields (representing errors in the forcing data) and is also subject to randomly generated noise that is added to the model parameters and prognostic variables (representing errors in model physics and parameters). The error covariance matrices that are required for the filter update can then be diagnosed from the spread of the ensemble at the update time. The EnKF is flexible in its treatment of errors in model dynamics and parameters. It is also very suitable for modestly nonlinear problems and has become a popular choice for land data assimilation [*Andreadis and Lettenmaier*, 2005; *Durand and Margulis*, 2008; *Kumar et al.*, 2007; *Pan and Wood*, 2006; *Reichle et al.*, 2002a, 2002b; *Zhou et al.*, 2006].

[10] The EnKF works sequentially by performing in turn a model forecast and a filter update. Formally, the forecast step for ensemble member i can be written as

$$x^-_{t,i} = f(x^+_{t-1,i}, q_{t,i}) \tag{1}$$

where $x^-_{t,i}$ and $x^+_{t-1,i}$ are the forecast (denoted with $^-$) and analysis (denoted with $^+$) state vectors at times t and t−1, respectively. The model error (or perturbation vector) is denoted with $q_{t,i}$ and its covariance with Q_t. The filter update produces the analyzed state vector $x^+_{t,i}$ at time t and can be written as

$$x^+_{t,i} = x^-_{t,i} + K_t\,(y_{t,i} - H_t\,x^-_{t,i}) \tag{2}$$

where $y_{t,i}$ denotes the observation vector (suitably perturbed) and H_t is the observation operator (which is assumed linear for ease of notation, although this requirement can be relaxed). The Kalman gain matrix K_t is given by

$$K_t = P_t\,H_t^T\,(H_t\,P_t\,H_t^T + R_t)^{-1} \tag{3}$$

where P_t is the forecast error covariance (diagnosed from the ensemble $x^-_{t,i}$), R_t is the observation error covariance, and superscript T denotes the matrix transpose. Simply put, the Kalman gain K_t represents the relative weights given to the model forecast and the observations based on their respective uncertainties, along with the error correlations between different elements of the state vector. If the system is linear, if its model and observation error characteristics satisfy certain assumptions (including Gaussian, white, and uncorrelated noise), and if the input error parameters are correctly specified, the Kalman gain of equation (3) is optimal in the sense of minimum estimation error variance. Unfortunately, these optimal conditions are never simultaneously met in land data assimilation applications. Our focus in this paper is on the accurate specification of the input error variances.

2.2. The Adaptive EnKF

[11] The central idea behind adaptive filtering methods is that internal diagnostics of the assimilation system should be consistent with the values that are expected from input parameters provided to the data assimilation system. The most commonly used diagnostics for adaptive filtering are based on the observation-minus-forecast residuals or innovations (computed here as $v_t \equiv E\{y_{t,i} - H_t\,x^-_{t,i}\}$, where $E\{\cdot\}$ is the ensemble mean operator). For a linear system operating under optimal conditions, the lagged innovations covariance is

$$E[v_t\,v_{t-k}^T] = \delta_{k,0}\,(H_t P_t H_t^T + R_t) \tag{4}$$

where $E[\cdot]$ is the expectation operator and $\delta_{k,0}$ is the Kronecker delta. Equation (4) implies that the innovations sequence is uncorrelated in time and that its covariance is equal to the sum of the forecast error covariance $H_t\,P_t\,H_t^T$

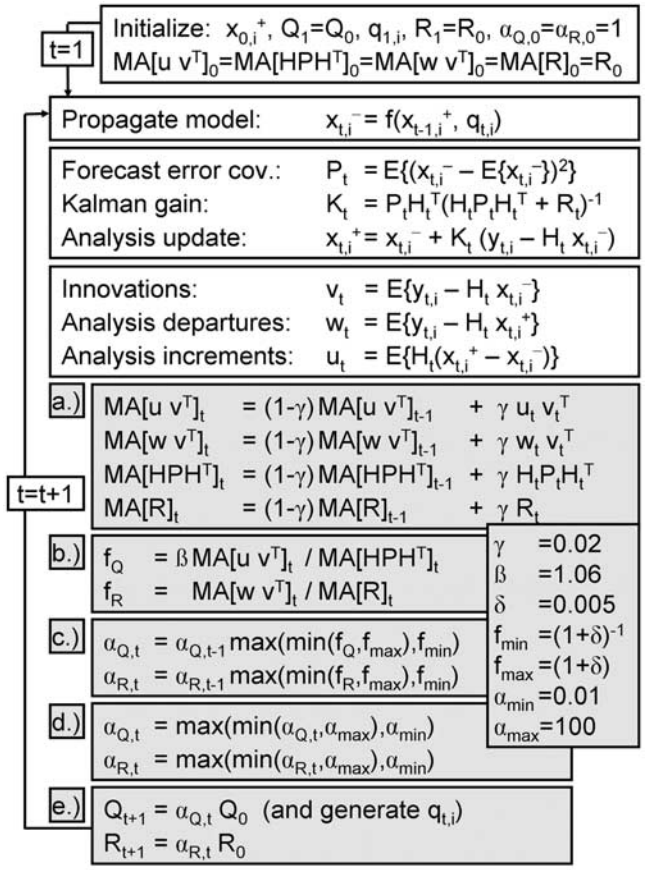

Figure 1. Adaptive EnKF algorithm. Initial guesses for model and observation error covariances are denoted with Q_0 and R_0, respectively. Shaded boxes indicate the adaptive module of the EnKF. "MA[·]" denotes exponential moving average, E{·} denotes ensemble mean. See text for symbols and further discussion.

(in observation space) and the observation error covariance R_t. Now recall that the forecast error covariance P depends on the model error covariance Q. If the innovations show less spread than expected, the input error covariances (Q and/or R) are too large, and vice versa. Such information can be used for adaptive tuning of Q and/or R.

[12] Alternative diagnostics are based on the analysis departures $w_t = E\{y_{t,i} - H_t x_{t,i}^+\}$ and the (observation space) analysis increments $u_t = E\{H_t(x_{t,i}^+ - x_{t,i}^-)\}$. For linear systems operating under optimal conditions we have [*Desroziers et al.*, 2005]

$$E[u_t v_t^T] = H_t P_t H_t^T \tag{5}$$

$$E[w_t v_t^T] = R_t \tag{6}$$

Equations (5) and (6) are attractive because they suggest a simple way of estimating the model and observation error covariances separately. This approach is the basis of the proof-of-concept example of *Desroziers et al.* [2005] and it is also the approach employed here. Good results with this approach have also been reported by other groups (Eugenia Kalnay, personal communication, 2007).

[13] A flowchart of the adaptive EnKF algorithm is provided in Figure 1. After appropriate initialization, the adaptive EnKF starts with a regular EnKF forecast and update step (equations (1)–(3)), along with the computation of the innovations, the analysis departures, and the analysis increments. The adaptive module follows thereafter. In our implementation, the adaptive filter is estimating the ratios of the true input error variances and their *initial* values, that is we are ultimately trying to determine (separately for each location) the scalar factors $\alpha_Q = \text{trace}(Q_{true})/\text{trace}(Q_{t=0})$ and $\alpha_R = \text{trace}(R_{true})/\text{trace}(R_{t=0})$, which will henceforth be referred to as "adaptive tuning factors." To this end, the adaptive module starts with estimating the terms in equations (5) and (6). In atmospheric data assimilation systems, error characteristics vary smoothly and at fairly large horizontal scales when compared to the size of the computational units. Therefore adaptive algorithms in atmospheric assimilation systems typically compute the terms of equations (5) and (6) by averaging across the domain (or large parts of it). Because of the substantial heterogeneity of the land surface, this approach is unlikely to work well for soil moisture data assimilation. Instead, for each catchment we compute the desired statistic from the relevant time series for that single catchment (and thereby reduce equations (5) and (6) to sets of scalar equations). The most recent estimate of $E[u\ v^T]$, for instance, is approximated as an exponential moving average, that is, at update time t we approximate $E[u\ v^T] \cong MA[u\ v^T]_t$, where the exponential moving average is denoted with MA[·] and defined as $MA[u\ v^T]_t \equiv (1-\gamma)\,MA[u\ v^T]_{t-1} + \gamma\ u_t\ v_t^T$, with an ad-hoc choice of $\gamma = 0.02$ for the experiments presented here. Similarly, we compute time-filtered estimates of $E[w\ v^T]$, $H\ P\ H^T$, and R (Figure 1a).

[14] Next, we determine the ratio of actual covariances and expected covariances, that is we compute the ratios $f_Q = \beta\ MA[u\ v^T]_t/MA[H\ P\ H^T]_t$ and $f_R = MA[w\ v^T]_t/MA[R]_t$ (Figure 1b; the factor β will be explained below). We then use these ratios to update the adaptive tuning factors α_Q and α_R based on the idea (expressed in equations (5) and (6)) that these ratios should be close to unity for correct input error parameters. Because the moving average estimates of the terms in equations (5) and (6) are very noisy even after temporal smoothing, it is necessary to further restrict the ratios f_Q and f_R to the interval $[(1+\delta)^{-1}, (1+\delta)]$, with $\delta = 0.005$ used here (Figure 1c). In practice, this means that α_Q and α_R are decreased or increased by only a very small fraction at each update time, no matter how far the original ratios f_Q and f_R are from unity. While slowing down convergence considerably, such a restriction is necessary for stabilizing the algorithm. In addition, we restrict α_Q and α_R to the interval [0.01, 100] (Figure 1d). Finally, we derive the latest estimates of the true Q and R by scaling the original input model parameters $Q_{t=0}$ and $R_{t=0}$ with the latest estimates of α_Q and α_R, respectively (Figure 1e). Note that for lognormally distributed, multiplicative perturbations (such as are applied to precipitation and shortwave radiation forcing, section 3.2), the covariance scaling is applied to the corresponding normal deviates. The "scaled" lognormal perturbations are computed by transforming the scaled normal deviates into lognormal space. Note also that multiplicative perturbations (as opposed to additive perturbations) exacerbate the noisy character of the ratios f_Q and f_R.

[15] It is important to note that in our EnKF-based system we are interested in scaling the model error covariance Q, and not the forecast error covariance P [e.g., *Dee*, 1995; *Desroziers et al.*, 2005]. Our approach therefore makes a linearity assumption between P and Q that is only valid approximately. In particular, we found that the ratio f_Q (Figure 1b) is systematically underestimated for perfect input error parameters unless an empirical factor β is introduced. This factor can be determined in a twin experiment that uses perfect input error parameters. If the factor β is omitted, the adaptive algorithm tends to produce slightly worse results (when compared to the non-adaptive algorithm) in the case of correct initial input error parameters. It is, however, highly unlikely that the initial guess of input error parameters is already perfect, and the adaptive filter does yield significant improvements for most choices of input error parameters even when the factor β is omitted. It must also be noted that theoretically, the adaptive algorithm relies on stationary conditions, and Q_{true} and R_{true} should not vary with time. Nevertheless, slow variations of Q_{true} and R_{true} on seasonal timescales and longer can be handled by the algorithm without significant loss of performance.

3. Experiment Description

[16] We use a twin experiment to test the impact of erroneously specified model and observation error parameters and the performance of our adaptive filter. First, we conduct an ensemble integration of the land surface model with a known set of model errors (perturbations) and without data assimilation. A randomly chosen member from this reference ensemble integration serves as the "truth". A second model integration, this time without perturbations, represents the best model estimate of the truth without the benefit of data assimilation and is called the "open loop" integration. Next, synthetic retrieval data sets are generated from the true surface soil moisture and are subsequently assimilated back into the land surface model with various sets of (intentionally wrong) input error parameters using both the standard EnKF and the adaptive EnKF. Finally, the resulting assimilation estimates are compared against the truth data set. In the remainder of this section, we provide more detailed descriptions of the land surface model, the synthetic retrievals, and the data assimilation parameters.

3.1. Land Surface Model and Observations

[17] Model soil moisture is obtained from integrations of the NASA Catchment Land Surface Model [hereinafter Catchment model; *Koster et al.*, 2000]. The Red-Arkansas river basin in the United States was chosen as the experiment domain because it exhibits a range of conditions that favor or prohibit soil moisture retrieval from space. The assimilation experiments cover the (almost) 19-year period from 1 October 1981 through 1 September 2000. The time step for the model integration is 20 min. The meteorological forcing data for the Catchment model were provided at hourly time steps and at 1 km resolution [*Sharif et al.*, 2007]. Forcing data from 1 April 1980 to 1 October 1981 were used to spin up the model through repeated integrations.

[18] The Catchment model's basic computational unit is the hydrological catchment (or watershed). The Red-Arkansas river basin is divided into 308 catchments (excluding inland water) with a median linear scale of 35 km (ranging from 12 km to 96 km). In each catchment, the vertical profile of soil moisture is determined by the equilibrium soil moisture profile from the surface to the water table (defined by a balance of gravity and capillary forces) and by two additional variables that describe deviations from the equilibrium profile in a 1 m root zone layer and in a 2 cm surface layer, respectively. Unlike traditional, layer-based models, the Catchment model includes an explicit treatment of the spatial variation of soil water and water table depth within each hydrological catchment based on the statistics of the catchment topography.

[19] Synthetic retrievals of 2 cm surface soil moisture were obtained once daily from the truth integration after applying a space-time mask based on the expected availability of retrievals from an L-band satellite sensor [*Crow et al.*, 2005]. Three separate synthetic retrieval data sets were simulated by adding white noise (with standard deviations of 0.02, 0.05, and 0.08 m^3m^{-3}) to mimic retrieval errors. Earlier studies found large differences between the temporal moments of satellite and model soil moisture [*Reichle and Koster*, 2004; *Drusch et al.*, 2005]. Such biases are avoided by design in our twin experiment. In applications of the adaptive algorithm to satellite observations, any biases will have to be addressed with suitable methods such as climatological rescaling [*Reichle and Koster*, 2004] or bias estimation [*De Lannoy et al.*, 2007].

3.2. Data Assimilation Setup

[20] The reference ("truth") model error parameters are based on our experience with the assimilation of soil moisture retrievals from the Advanced Scanning Multichannel Radiometer for the Earth Observing System (AMSR-E) [*Reichle et al.*, 2007] and are summarized in Table 1. Zero-mean, normally distributed additive perturbations were applied to the downward longwave radiation forcing and to two Catchment model prognostic variables related to soil moisture (catchment deficit and surface excess). Lognormally distributed multiplicative perturbations (with mean 1) were applied to the precipitation and the downward shortwave radiation forcing. Time series correlations were imposed via a first-order auto-regressive model (AR(1)) for all perturbation fields. In this study we used a one-dimensional EnKF [*Reichle and Koster*, 2003] and therefore did not use spatially correlated perturbation fields. Cross-correlations were only imposed on perturbations of the precipitation and radiation fields.

[21] A randomly chosen member from an ensemble integration with the reference model error parameters (Table 1) serves as the synthetic truth. The open loop integration (no perturbations, no assimilation) has a surface soil moisture RMSE of 0.035 m^3m^{-3} (volumetric) and a bias of -0.006 m^3m^{-3}. Throughout the manuscript, biases and RMS errors are computed relative to the truth data and averaged over the 19-year experiment period and the entire domain after applying the same space-time mask to the assimilation estimates that was used in the generation of the synthetic retrievals. The mask is used to discount the impact of times and locations when data were not assimilated with the one-dimensional filter employed here.

[22] Starting from the true model error parameters of Table 1, we defined five additional sets of input model error parameters for use in the assimilation integrations

Table 1. Parameters for Perturbations to Meteorological Forcing Inputs and Catchment Model Prognostic Variables in the Truth Integration

Perturbation	Additive (A) or Multiplicative (M)?	Standard Deviation	AR(1) Time Series Correlation Scale	Cross-Correlation With Perturbations in SW	Cross-Correlation With Perturbations in LW
Precipitation	M	0.5	1 day	−0.8	0.5
Downward shortwave (SW)	M	0.3	1 day	n/a	−0.5
Downward longwave (LW)	A	50 W m^{-2}	1 day	n/a	n/a
Catchment deficit	A	0.05 mm	12 h	n/a	n/a
Surface excess	A	0.02 mm	12 h	n/a	n/a

(Table 2). The six sets of model error parameters differ only in the strength of the perturbations (model error variances) and otherwise share the same characteristics in terms of the sources of error (which forcing fields and model states are uncertain) and in the temporal and cross-correlation structure. Moreover, the relative strength of perturbations to each field within all parameter sets is similar (though not identical). Using the different sets of model error parameters in 12-member ensemble integrations (without assimilation) leads to ensemble spreads in surface soil moisture ranging from 0.005 m^3m^{-3} to 0.093 m^3m^{-3} (Table 3). By design, the ensemble spread underestimates the actual error of the open loop integration (0.035 m^3m^{-3}) for model error parameter sets A and B and overestimates it for parameter sets D, E, and F.

[23] For comparison, the RMS errors of the ensemble mean surface soil moisture from the ensemble integrations (without assimilation) range from 0.030 m^3m^{-3} to 0.099 m^3m^{-3} (Table 3). Ideally, the actual errors of all ensemble integrations would match the open loop error (0.035 m^3m^{-3}) because the perturbations are designed to be zero-mean to the extent possible given system non-linearities and computational constraints. This is the case (to within 0.005 m^3m^{-3}) for parameter sets A, B, C, and D. For parameter sets E and F, the perturbations are so strong that biases become significant (Table 3). Such biases are an unavoidable consequence of non-linearities in the system, such as the upper and lower bounds of soil moisture. Note that the ensemble integration with the true model error parameter set C performs slightly better than the open loop. Its ensemble mean could be used instead of the open loop integration. The differences are minor, though, and do not affect our results and general conclusions.

[24] Next, we define five input observation error standard deviations (0.001; 0.02; 0.05; 0.08; and 0.11 m^3m^{-3}) that are used in combination with the six sets of input model error parameters (Table 2) in a suite of data assimilation experiments. For the retrieval data set with a true observation error standard deviation of 0.05 m^3m^{-3} (section 3.3) we performed 30 data assimilation experiments with the non-adaptive EnKF (and another 30 with the adaptive EnKF) by combining each of the five input observation error standard deviations (listed above) with each of the six sets of input model error parameters listed in Table 2. These experiments will be the focus of the remainder of the paper. Each of these 60 data assimilation experiments is over the 19-year experiment period and the entire Red-Arkansas domain and uses 12 ensemble members. We also performed an additional suite of 80 experiments in which we assimilated the retrieval data sets with true observation error standard deviations of 0.02 and 0.08 m^3m^{-3} using select combinations of the input model and observation error parameters. The results from these additional experiments (not shown) are qualitatively very similar and support the same general conclusions.

4. Results

4.1. Impact of Input Error Parameters on Assimilation Estimates

[25] Each data assimilation experiment with the standard EnKF has a unique set of input error parameters (Table 2 and section 3) that leads to a unique pair of scalars: the (space and time) average forecast error variance (P_0) and the input observation error variance (R_0). We can thus plot two-dimensional surfaces of filter performance as a function of sqrt(P_0) and sqrt(R_0). Figure 2a, for example, shows one such surface with the performance measure being the RMSE of surface soil moisture estimates from the (non-adaptive) EnKF. Each of the 30 plus signs in the figure indicates the performance of a 19-year assimilation integration over the entire Red-Arkansas domain. For example, the experiments that use the true model error parameters (set "C" in Table 2) have a forecast error standard deviation of about 0.02 m^3m^{-3} in data assimilation mode. (In each assimilation experiment, the value of the forecast error

Table 2. Standard Deviations for Perturbations to Meteorological Forcing Inputs and Catchment Model Prognostic Variables in the Assimilation Integrations[a]

	A	B	C ("true")	D	E	F
Precipitation	0.05	0.2	0.5	0.75	0.875	1.0
Downward shortwave	0.03	0.1	0.3	0.5	0.75	1.0
Downward longwave	5 W m^{-2}	25 W m^{-2}	50 W m^{-2}	100 W m^{-2}	150 W m^{-2}	200 W m^{-2}
Catchment deficit	0.005 mm	0.02 mm	0.05 mm	0.1 mm	0.3 mm	0.5 mm
Surface excess	0.002 mm	0.01 mm	0.02 mm	0.04 mm	0.12 mm	0.2 mm

[a]All other parameters as in Table 1. Parameter set C is identical to the true error parameters of Table 1.

Table 3. Ensemble Spread (Ensemble Standard Deviation Averaged Over Experiment Period and Domain), RMSE, and Bias of Surface Soil Moisture for the Open Loop (No Perturbations) and for Ensemble Integrations (No Assimilation) With the Model Error Parameter Sets A Through F Defined in Table 2[a]

	Open loop	A	B	C ("true")	D	E	F
Ensemble spread	n/a	0.005	0.017	0.031	0.045	0.077	0.093
RMSE	0.035	0.034	0.032	0.030	0.038	0.075	0.099
Bias	−0.006	−0.006	−0.006	−0.001	0.017	0.050	0.065

[a]Units are m^3m^{-3}.

variance depends primarily on the input model error parameter set (Table 2) and varies slightly with the input observation error variance. Note that the ensemble spread listed in Table 3 is without data assimilation.)

[26] Figure 2a illustrates that the estimation error in surface soil moisture is smallest near the experiment that uses the true model and observation error inputs. The minimum estimation error is around 0.02 m^3m^{-3}, down from the open loop value of 0.035 m^3m^{-3}. The estimation error increases as the input error parameters deviate from their true values. Generally, the estimation errors in surface soil moisture are quite sensitive to both the assumed model and the assumed observation error variances. Figure 2a also indicates where the estimation error surface intersects the

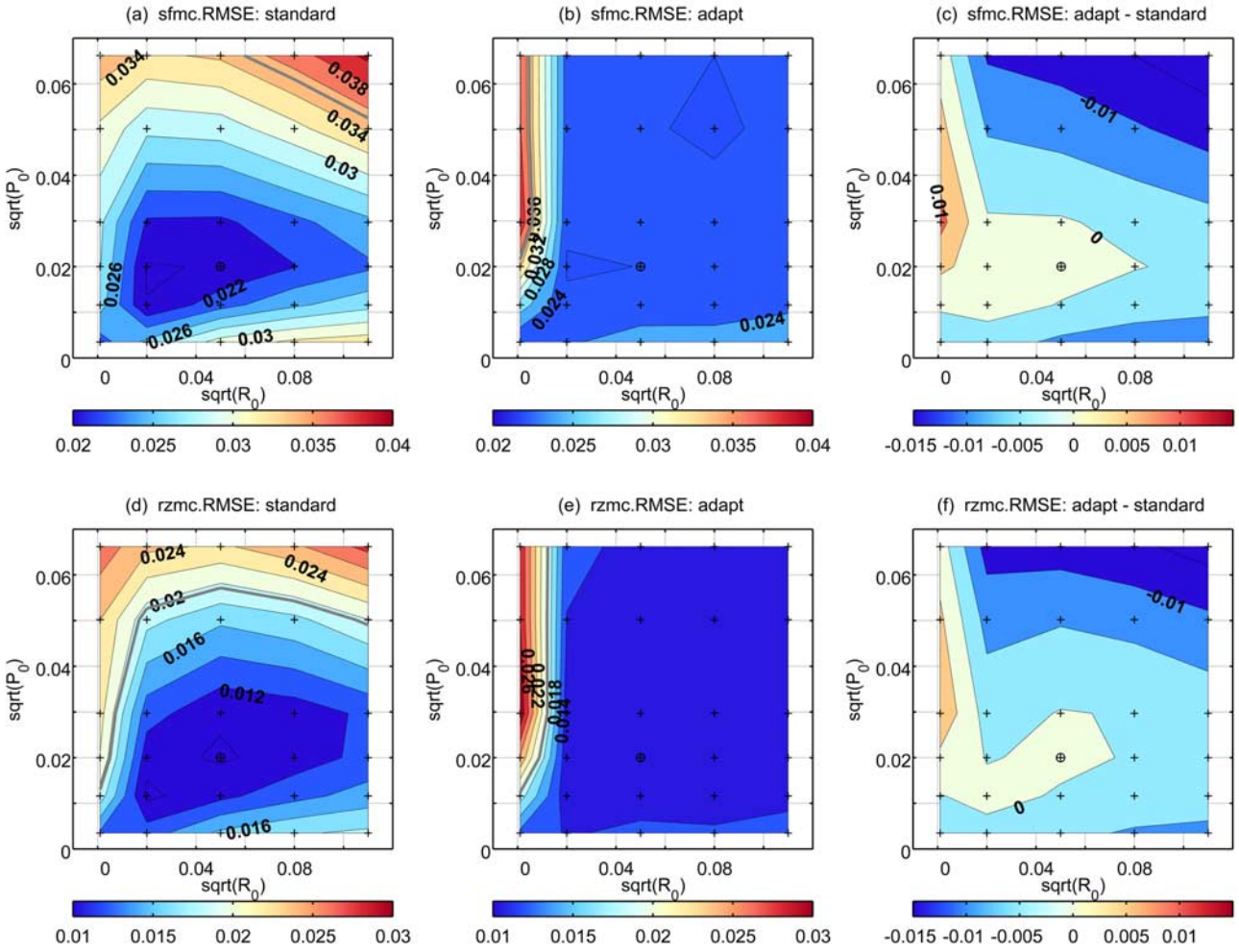

Figure 2. RMSE of standard EnKF estimates for (a) surface and (d) root zone moisture as a function of the (ordinate) time and space average forecast and (abscissa) input observation error standard deviations of surface soil moisture. Similarly, RMSE of adaptive EnKF estimates for (b) surface and (e) root zone soil moisture. RMSE difference plots (adaptive minus standard EnKF) for (c) surface and (f) root zone soil moisture. Units are volumetric percent (m^3 m^{-3}). Each plus sign indicates the result of a 19-year assimilation integration over the entire Red-Arkansas domain. The circled plus sign indicates the experiment that uses the true error parameters in the data assimilation system. Thick gray lines indicate RMSE of open loop (no assimilation) integration.

open loop error. For grossly overestimated model and observation error variances, the assimilation estimates of surface soil moisture are in fact slightly worse than the open loop estimates.

[27] Figure 2d shows the same for the estimation error in root zone soil moisture. Qualitatively, the result is similar to that for surface soil moisture, that is, (i) the minimum estimation error ($\sim$0.01 m^3m^{-3}) is near the experiment with the true input error parameters, (ii) the estimation error gradually increases as the input error parameters deviate from their true values, and (iii) there is sensitivity to both the input model and observation error variances. For root zone soil moisture, however, the assimilation estimates are somewhat worse than the open loop error of 0.02 m^3m^{-3} not only for gross overestimation of the input error variances but also for severe underestimation of the observation error variance.

[28] Figures 3a, 3d, and 3g show the performance of the non-adaptive EnKF for the latent heat flux, the sensible heat flux, and the total runoff, respectively. Similar to the estimation errors in soil moisture, the estimation errors in the fluxes are smallest near the experiment that uses the true input error parameters. The assimilation estimates, however, improve only modestly over the open loop integration. The minimum estimation errors are 24 Wm^{-2}, 19 Wm^{-2}, and 0.28 mm d^{-1} for latent heat flux, sensible heat flux, and runoff, respectively, compared with open loop errors of 27 Wm^{-2}, 22 Wm^{-2}, and 0.31 mm d^{-1}, respectively. Poorly specified input error parameters easily lead to assimilation estimates of fluxes that are worse than the model estimates without data assimilation. This is particularly true for runoff estimates. For overestimated model error variances and in particular for severely underestimated observation error variance, the runoff estimates get dramatically worse.

[29] The filter performance for the surface heat fluxes is fairly insensitive to the assumed observation error variance but varies strongly with the assumed model error variance because the relatively small benefit bestowed on the flux estimates through the assimilation of surface soil moisture retrievals is canceled out by the biases that are introduced through the unavoidable non-linearities in the system. In other words, inherent non-linearities in the ensemble generation explain most of the sensitivity of the flux estimates to the input model error variance.

[30] Finally, we evaluate the performance of the EnKF in terms of its internal diagnostics and how these diagnostics depart from their optimal values (in a linear system) or correct values (with respect to the truth). We generally refer to these differences as ''misfits''. For the ''innovations misfit'' we first compute normalized innovations by dividing the filter innovations with their expected standard deviation (see equation (4)). The ''innovations misfit'' is then defined as the spatial RMS difference between the (temporal) sample variance of the normalized innovations and its expected value of 1. We also evaluate the filter's estimates of the observation and analysis error standard deviations. The former is obviously specified as an input parameter (and adjusted in the adaptive filter). The latter is simply the ensemble standard deviation of surface soil moisture after the Kalman update. Both should approximate the actual RMS errors. We measure errors in these uncertainty estimates as the RMS difference between the filter's uncertainty estimates and the actual RMS errors (with respect to the truth) of the observations and the assimilation estimates, respectively.

[31] Figure 4a shows the logarithm (base 10) of the innovations misfit for the non-adaptive EnKF experiments similar to the surface plots of Figures 2 and 3. The minimum innovations misfit clearly coincides with the optimal model and observation error parameters. The innovations misfit for the corresponding experiment is 0.05(= $10^{-1.3}$) (dimensionless). Extreme values of the innovations misfit are seen when the observation error standard deviation is severely underestimated. Generally, the innovations misfit is sensitive to the input model and observation error parameters. Figure 4d (trivially) shows the (designed) error in the filter's estimate of the observation error standard deviation. The error in the filter's estimate of the output uncertainty (or analysis error) in surface soil moisture is shown in Figure 4g. For input error parameters close to the true values, the filter estimate of the analysis error standard deviation is within 0.003 m^3 m^{-3} of the actual RMS error. For severely underestimated model or observation error standard deviations the filter produces very poor estimates of the uncertainty in its output that are as far away as 0.03 m^3 m^{-3} from the actual RMS errors. (Note that the misfit in the analysis error for each assimilation experiment is computed from the actual errors for that experiment.) Figure 4g indicates that poorly specified model error parameters lead not only to poor estimates of the fields themselves but also to poor estimates of their uncertainty.

4.2. Adaptive Filtering

[32] As described above, all data assimilation experiments were also conducted with the adaptive EnKF. Before discussing the performance of the adaptive filter in terms of RMS errors, it is instructive to examine the adaptive tuning factors α_Q and α_R that are estimated by the adaptive filter. Figure 5 shows that the estimated α_Q and α_R converge to stable values. Since in our synthetic experiment we know the true input error parameters, we can determine the ''true'' α_Q and α_R for each case (also shown in Figure 5). Generally, the adaptive algorithm is able to identify α_R quite successfully. Only in the case of severely underestimated observation error does the estimated α_R not converge toward the true value of 2500 (outside the plotted range in Figure 5), because α_R is limited by the imposed cap (section 2). Similarly, α_Q is identified adequately by the adaptive algorithm, albeit with less skill in the case of severely over- or underestimated model error variance, where α_Q is again limited by the imposed cap. Increasing (or even eliminating) these caps would presumably lead to improved results in the extreme cases, but might lead to unforeseen problems in other situations.

[33] The estimates of α_Q and α_R vary across the domain (Figure 5), even though the experiment setup (with homogeneous true and assumed input error parameters) suggests that they should not. The spatial variability is particularly pronounced for estimates of α_Q. There is no indication that the spatial patterns of the α_Q estimates (not shown) are directly related to variations in soil type, vegetation characteristics, or the climate across the domain. One reason for the relatively greater difficulty in estimating the model error variance is the non-linear relationship between model errors (for example in precipitation or surface excess) and errors in

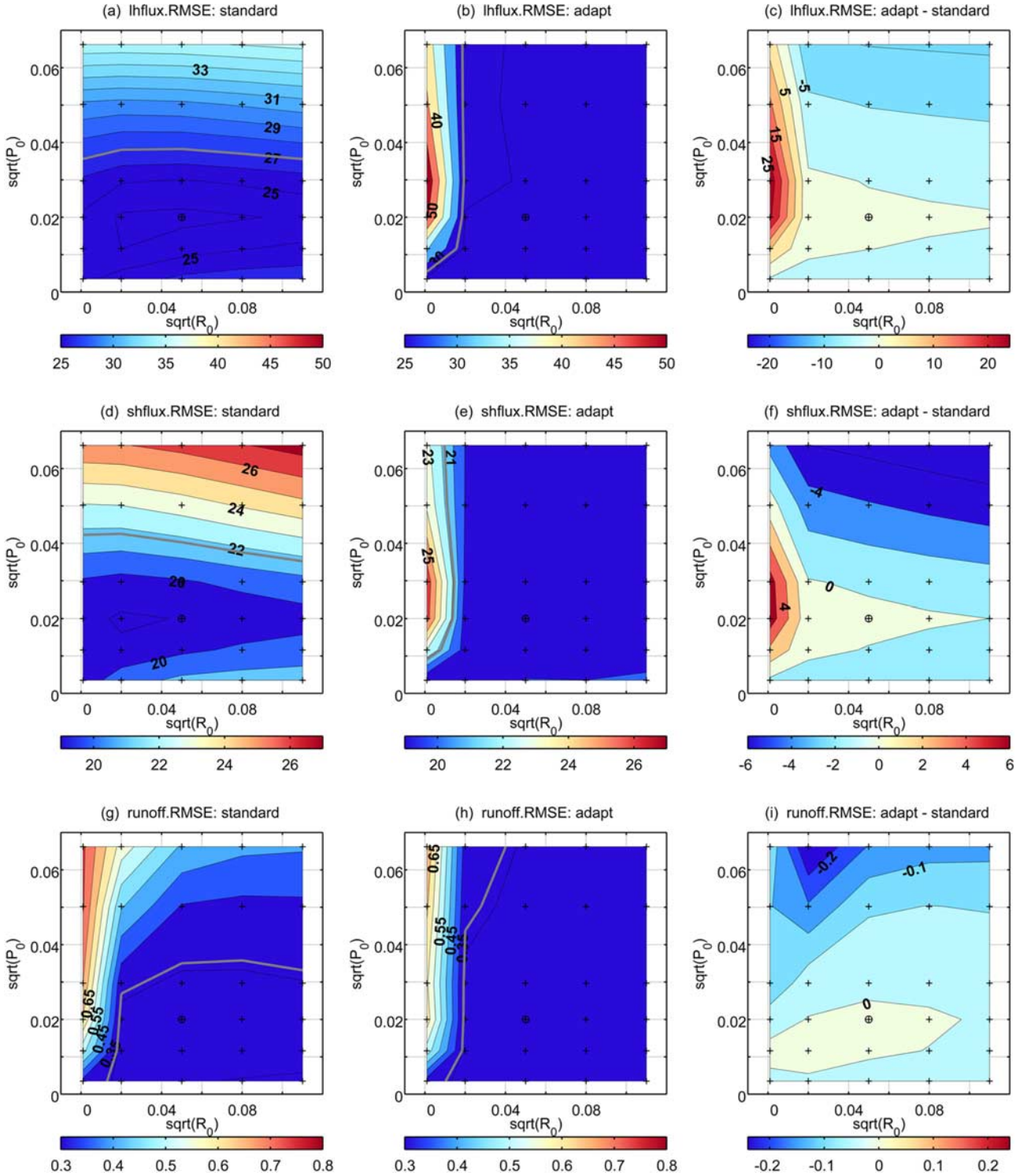

Figure 3. RMSE of standard EnKF estimates for (a) latent heat flux, (d) sensible heat flux, and (g) runoff. RMSE of adaptive EnKF estimates for (b) latent heat flux, (e) sensible heat flux, and (h) runoff. RMSE difference plots (adaptive minus standard EnKF) for (c) latent heat flux, (f) sensible heat flux, and (i) runoff. Latent and sensible heat fluxes are in units of (W m^{-2}), runoff is in (mm d^{-1}). Ordinate, abscissa, plus signs, and gray lines as in Figure 2.

the observed surface soil moisture. Moreover, the model error characteristics used here (motivated by our experience with satellite data assimilation) include non-linear (multiplicative) model errors with temporal correlations that technically violate the assumptions of the EnKF. Such violations contribute to the high level of noise in the quantities that are used for estimating the adaptive tuning factor α_Q for the model error variance. Because the EnKF

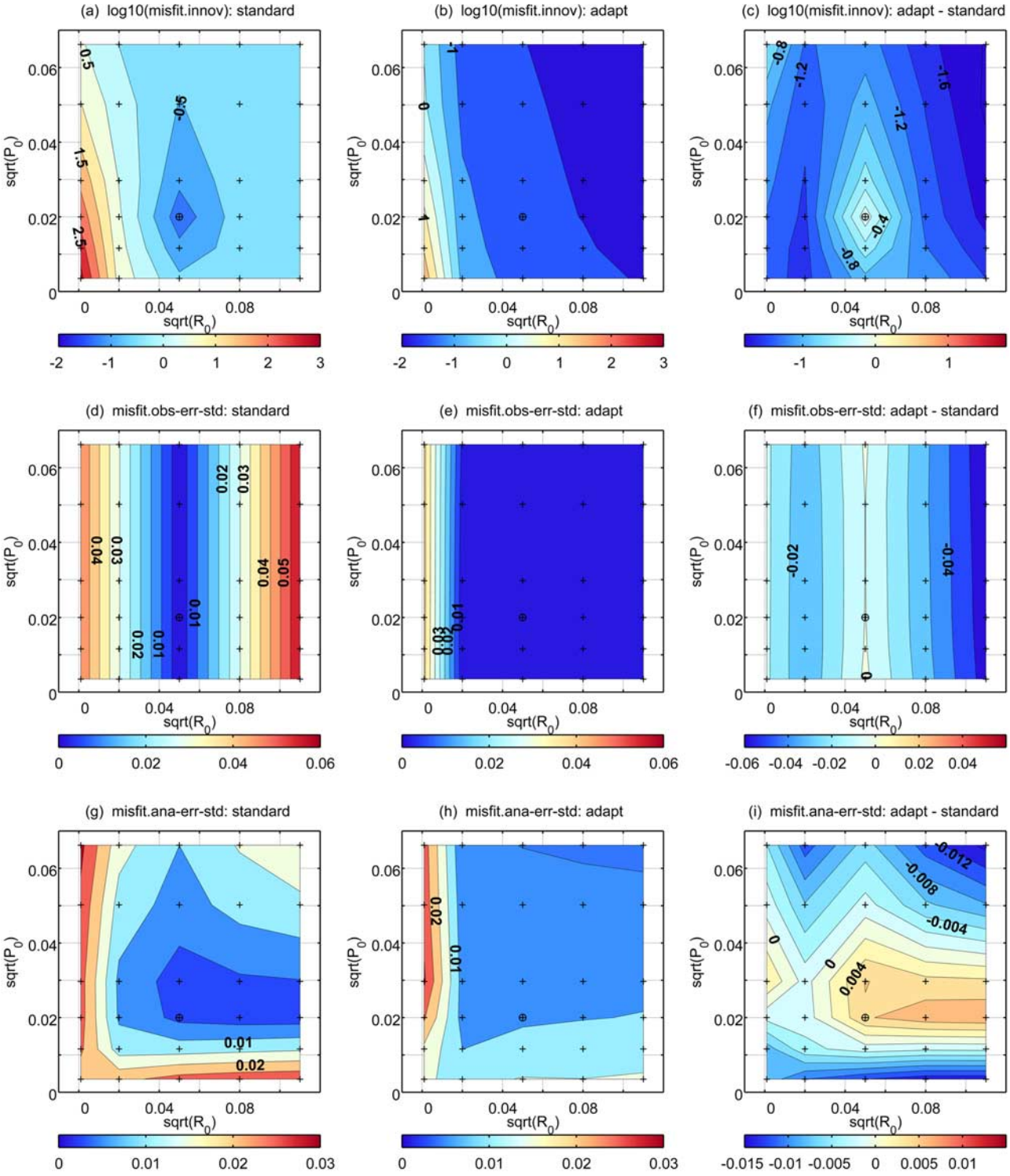

Figure 4. Diagnostics of the standard EnKF: (a) log-10 of the normalized innovations misfit (dimensionless), (d) the error in the filter estimate of the observation error standard deviation (m^3m^{-3}), and (g) the error in the filter estimate of the analysis error standard deviation (m^3m^{-3}). Same diagnostics but for the adaptive EnKF in (b), (e), and (h), respectively. Difference plots for same diagnostics (adaptive minus standard EnKF) in (c), (f), and (i), respectively. Ordinate, abscisssa, and plus signs as in Figure 2.

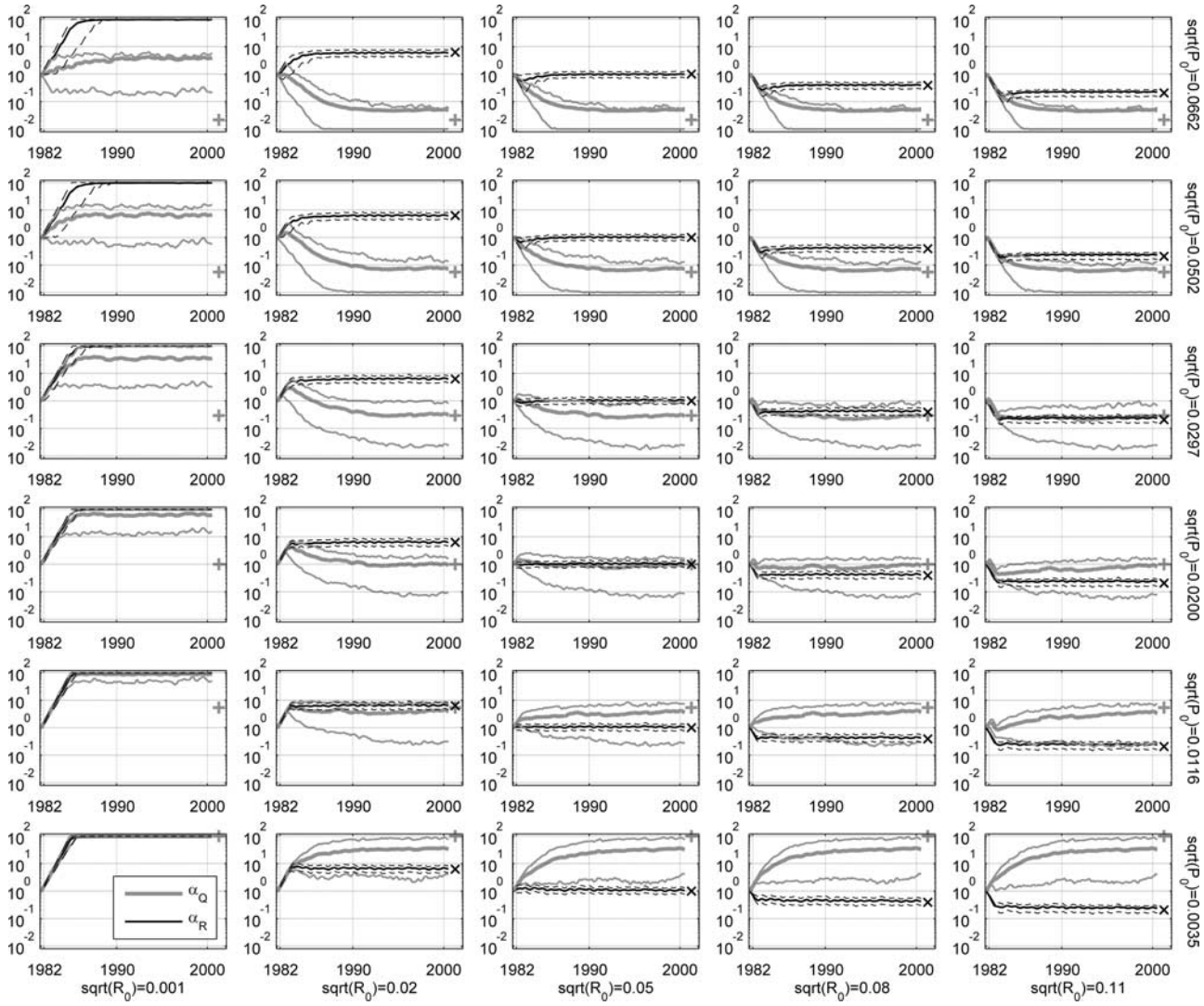

Figure 5. Monthly and area average estimates of (thick gray line) α_Q and (solid black line) α_R from the adaptive EnKF. Thin gray and dashed black lines show (spatial) 10th and 90th percentiles of α_Q and α_R, respectively. Gray plus signs and black crosses indicate ''true'' values of α_Q and α_R, respectively (see text).

assumptions are less severely violated in the case of the observation error characteristics, it is generally easier to estimate the observation error variance.

[34] The convergence timescales for α_Q and α_R are typically on the order of a few years and are primarily controlled by the parameter δ (Figure 1b). A larger value for δ generally improves the speed of convergence but also leads to noisier and oscillating estimates of α_Q and α_R later on. Here, we use an empirical value of $\delta = 0.005$. Future studies may show how δ is related to the time it takes to obtain robust error statistics, most importantly of precipitation errors. Note also that generally, α_R converges faster than α_Q for the same reasons that estimates of α_R are less spatially variable than estimates of α_Q; observation error characteristics are generally better behaved than model error characteristics when it comes to violating EnKF assumptions. The slow convergence speed of the adaptive scheme is clearly an undesirable feature and is hardly acceptable in practice. A simple and affordable solution to the problem is to assimilate a given data set into a given model several times and each time use the final estimates of the input error parameters as initial input error parameters in the next assimilation integration. It should also be noted that in operational systems even minor improvements, as are realized before the adaptive filter converges, are valuable.

[35] Finally, it is interesting to note that during the initial phase the adaptive algorithm appears to optimize the sum of the model and observation error variances, and only later distinguishes between the two components. This can be seen in Figure 5, for example, in the panel for $\text{sqrt}(R_0) = 0.02$ and $\text{sqrt}(P_0) = 0.0297$ (third row, second column). In this case, the input model error variances are overestimated (with respect to the truth) while the observation error variance is underestimated to a relatively larger degree. The adaptive algorithm at first increases both the model and the observation error variances until the observation error variance is close to its true value. Only then does the algorithm reduce the model error variance.

[36] Let us now revisit Figures 2, 3, and 4 and examine the skill of the adaptive filter estimates in terms of RMS errors in the land surface fields. Again, the adaptive filter uses evolving model and observation error variances. When drawing the surface plots for the adaptive EnKF, however, we use the forecast and observation error standard deviations that correspond to the *initial* input error parameters (and match those in the non-adaptive EnKF) on the ordinate and the abscissa, respectively.

[37] Figures 2b and 2e show estimation errors in surface and root zone soil moisture for the adaptive EnKF, and Figures 2c and 2f show the corresponding RMSE difference plots between the adaptive and the non-adaptive EnKF. It is immediately obvious that the adaptive EnKF performs better than or similar to the non-adaptive EnKF (in terms of soil moisture) across almost the entire range of initial model and observation error parameters. The improvements facilitated by the adaptive tuning are generally commensurate with the errors in the initial input model and observation error variances and lead to homogeneous skill in the assimilation estimates, except for severely underestimated observation error variance. In this case, the adaptive filter performs worse than the non-adaptive version because the cap on α_R leads to poor estimates of model error variances (Figure 5). Starting the adaptive filter with a conservative (that is, larger) initial estimate of the observation error variance offers an obvious solution that avoids the issue.

[38] Figures 3b, 3e, and 3h show estimation errors in the latent heat flux, the sensible heat flux, and the total runoff for the adaptive EnKF. Figures 3c, 3f, and 3i show the corresponding RMSE differences between the adaptive and the non-adaptive EnKF. The skill in flux estimates produced by the adaptive filter is qualitatively similar to the skill in soil moisture estimates. Generally, the adaptive tuning leads to smaller errors in latent heat flux, sensible heat flux, and runoff. The strong sensitivity of the estimation error to the input model error variance is substantially decreased. As for soil moisture, severe underestimation of the observation error variance leads to substantially worse flux estimates. Again, initializing the adaptive filter with a conservative estimate of the observation error variance should avoid the issue.

[39] Next, Figures 4b, 4e, and 4h assess the internal filter diagnostics of the adaptive filter. By design, the innovations misfit is greatly reduced for all data assimilation experiments by the adaptive filter (Figure 4b). More interestingly, the adaptive filter's estimates of the observation error standard deviations correspond far more closely to the actual RMS errors of the observations than those of the non-adaptive filter (Figure 4e). Improvements are seen even if the input observation error standard deviation is much too small. Finally, the adaptive filter produces better estimates of the analysis error in surface soil moisture than the non-adaptive filter when the input model error standard deviations are either substantially too small or substantially too large (Figure 4h). If, however, the initial input model error variance is close to its true value, the adaptive filter produces uncertainty estimates that are less reliable than those of the non-adaptive EnKF. These results are consistent with the difference in skill in the estimation of α_Q and α_R (Figure 5). Generally, the adaptive filter estimates the observation error variance quite successfully, but the estimation of the model error and (consequently) analysis error variances is more difficult.

[40] We tried several alternative adaptive tuning schemes, but all variants performed no better than the version described above. Among the alternatives were (i) an algorithm based solely on the innovations covariance (equation (4) with k = 0) and (ii) an algorithm based on the innovations covariance and lag-1 correlation (equation (4) with k = 0 and k = 1). Our approach for implementing the alternative algorithms was similar to the one described above, that is the alternative algorithms were also based on running estimates of the relevant diagnostics and on nudging of the observation and model error parameters such that the diagnostics are closer to their ideal values. Note that with approach (ii) the temporal decorrelation constraint (equation (4) with k = 1) only implies a direction in which observation and model error variances may be tuned [*Crow and Bolten*, 2007], but it is not clear by how much. For linear systems a closed-form solution exists to estimate the exact input error parameters [*Mehra*, 1970]. We conducted preliminary tests with this closed-form algorithm in a simpler land modeling system with good results (not shown). It is, however, difficult to implement such a closed-form solution in our quasi-operational system, and our choice of algorithm is motivated partly by the ease with which it can be implemented. Note also that in our system the innovations may not fully de-correlate anyway because of temporal error correlations in the forcing fields.

5. Summary and Conclusions

[41] Input error parameters to land surface data assimilation systems are difficult to determine and inherently uncertain. Wrongly specified input error parameters typically lead to degraded assimilation estimates that may even be worse than the model estimates or the observations alone. Adaptive filtering methods have been developed that try to estimate the model and observation error parameters and thus improve the assimilation estimates. In this paper, we investigated the impact of the input error parameters on the performance of a soil moisture data assimilation system in a 19-year twin experiment over the Red-Arkansas river basin. We also implemented an adaptive EnKF (Figure 1) and tested its performance.

[42] Our experiments clearly show that poorly specified input error parameters negatively affect the assimilation estimates of soil moisture, the surface fluxes, and runoff. For soil moisture, a non-adaptive EnKF generally performs well, even when obviously wrong input error parameters are used, and produces assimilation estimates that are an improvement over the open loop (no assimilation) estimates (Figure 2). The assimilation estimates of soil moisture are worse than the open loop estimates only when the model error variance is severely overestimated or the observation error variance is severely underestimated. Adaptive tuning with our algorithm generally improves the soil moisture estimates produced by the assimilation system, except when the observation error variances are underestimated by several orders of magnitude. Conservatively overestimating the observation error standard deviation at the start of the adaptive filter integration should avoid this problem.

[43] Poorly specified input error parameters have a strong impact on estimates of surface heat fluxes and runoff, primarily because the benefit of assimilating surface soil moisture observations for the estimation of these fluxes is modest even with perfect input error parameters (Figure 3). When the model error variance is overestimated, the large perturbations result in biases (because of unavoidable non-linearities in the system) that negatively impact the flux estimates. The adaptive filter greatly alleviates the problem and produces generally improved flux estimates. Finally, the adaptive filter is quite successful in estimating the true observation error variance (Figure 4). In the case of substantially over- or under-estimated model error variances, the adaptive filter also produces improved uncertainty estimates (for its surface soil moisture output) when compared to the non-adaptive filter, but the analysis uncertainty estimates from the adaptive filter are somewhat worse when the initial choices for the model error variances are close to the true error variances.

[44] The convergence timescales of the adaptive algorithm are typically on the order of a few years and require long data sets or repeated assimilation integrations before stable estimates of model and observation error variances can be reached (Figure 5). In operational systems, however, even minor improvements, as are realized before the adaptive filter converges, are valuable.

[45] The experiments described here provide many insights into the importance of choosing the correct input error parameters for soil moisture data assimilation, in particular in ensemble-based assimilation systems where biases are easily introduced unwittingly. The experiments were set up under favorable conditions. Our twin experiments used the same land surface model for generating the synthetic truth and for the assimilation. The filters, both the adaptive and the non-adaptive EnKF, were provided with perfect knowledge of which model states and forcing fields are the main source of uncertainty. Temporal and cross-correlation parameters were assumed perfectly known, and the relative strength of the errors within each model error parameter set was fairly similar. The remaining two main degrees of freedom were the error variances in the soil moisture observations and the corresponding model estimates. Even so, the adaptive filter did not perform optimally or flawlessly because of non-linearities inherent in the system and because computational constraints impose limits on the implementation of the adaptive tuning scheme.

[46] In future, the adaptive filter could be improved through further tuning of its parameters (in particular, β, γ, and δ of Figure 1). Moreover, the current adaptive algorithm produces model error variance estimates with a high degree of spatial variability. Localized spatial averaging of the diagnostics within the adaptive algorithm can perhaps be used to improve the performance of the filter. Including additional constraints such as the temporal decorrelation of the innovations sequence (equation (4) with $k > 0$) might also help, but it is not clear how this can be done in our system. The algorithm will also need to be tested with satellite retrievals of soil moisture rather than in a perfectly controlled twin experiment. Nevertheless, the experiments described here demonstrate the significant potential benefits of adaptive filtering for land data assimilation and are a necessary initial step into an area of land data assimilation research that has received little attention to date.

[47] **Acknowledgments.** We thank Hatim Sharif for providing the Red-Arkansas forcing data set and are grateful to Ming Pan and Randy Koster for helpful discussions. Computing was supported by the NASA High End Computing Program.

References

Anderson, J. L. (2007), An adaptive covariance inflation error correction algorithm for ensemble filters, *Tellus, Ser. A*, *59*(2), 210–224, doi:10.1111/j.1600-0870.2006.00216.x.

Andreadis, K., and D. Lettenmaier (2005), Assimilating remotely sensed snow observations into a macroscale hydrology model, *Adv. Water Resour.*, *29*, 872–886.

Bosilovich, M., J. Radakovich, A. da Silva, R. Todling, and F. Verter (2007), Skin temperature analysis and bias correction in a coupled land-atmosphere data assimilation system, *J. Meteorol. Soc. Jpn.*, *85A*, 205–228.

Crow, W., and J. Bolten (2007), Estimating precipitation errors using spaceborne surface soil moisture retrievals, *Geophys. Res. Lett.*, *34*, L08403, doi:10.1029/2007GL029450.

Crow, W. T., and E. Van Loon (2006), Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture, *J. Hydrometeorol.*, *7*, 421–432.

Crow, W. T., R. D. Koster, R. H. Reichle, and H. O. Sharif (2005), Relevance of time-varying and time-invariant retrieval error sources on the utility of spaceborne soil moisture products, *Geophys. Res. Lett.*, *32*(24), L24405, doi:10.1029/2005GL024889.

De Lannoy, G. J. M., R. H. Reichle, P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest (2007), Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter, *Water Resour. Res.*, *43*, W09410, doi:10.1029/2006WR005449.

Dee, D. P. (1995), On-line estimation of error covariance parameters for atmospheric data assimilation, *Mon. Weather Rev.*, *123*, 1128–1145.

Dee, D. P., and A. M. da Silva (1999), Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology, *Mon. Weather Rev.*, *127*(8), 1822–1834.

Dee, D., S. Cohn, A. Dalcher, and M. Ghil (1985), An efficient algorithm for estimating noise covariances in distributed systems, *IEEE Trans. Autom. Control*, *30*(11), 1057–1065.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005), Diagnosis of observation, background and analysis-error statistics in observation space, *Q. J. R. Meteorol. Soc.*, *131*(613), 3385–3396, Part C, doi:10.1256/qj.05.108.

Dong, J., J. P. Walker, P. R. Houser, and C. Sun (2007), Scanning multichannel microwave radiometer snow water equivalent assimilation, *J. Geophys. Res.*, *112*, D07108, doi:10.1029/2006JD007209.

Drusch, M. (2007), Initializing numerical weather prediction models with satellite derived surface soil moisture: Data assimilation experiments with ECMWF's integrated forecast system and the TMI soil moisture data set, *J. Geophys. Res.*, *112*(D3), D03102, doi:10.1029/2006JD007478.

Drusch, M., E. F. Wood, and H. Gao (2005), Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture, *Geophys. Res. Lett.*, *32*, L15403, doi:10.1029/2005GL023623.

Durand, M., and S. Margulis (2008), Effects of uncertainty magnitude and accuracy on assimilation of multiscale measurements for snowpack characterization, *J. Geophys. Res.*, *113*, D02105, doi:10.1029/2007JD008662.

Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, *53*, 343–367, doi:10.1007/s10236-003-0036-9.

Gelb, A. (Ed.) (1974), *Applied Optimal Estimation*, 374 pp., The MIT Press, Cambridge, MA.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model, 1: Model structure, *J. Geophys. Res.*, *105*(20), 24809–24822.

Kumar, S. V., R. H. Reichle, C. D. Peters-Lidard, R. D. Koster, X. Zhan, W. T. Crow, J. B. Eylander, and P. R. Houser (2007), A land surface data assimilation framework using the land information system: Description and applications, *Adv. Water Resour.*, doi:10.1016/j.advwatres.2008.01.013, in press.

Mehra, R. K. (1970), On identification of variances and adaptive Kalman filtering, *IEEE Trans. Autom. Control*, *AC15*(2), 175–184.

Mitchell, H. L., and P. L. Houtekamer (2000), An adaptive ensemble Kalman filter, *Mon. Weather Rev.*, *128*(2), 416–433.

Moghaddamjoo, A. R., and R. L. Kirlin (1993), Robust Adaptive Kalman Filtering, in *Approximate Kalman Filtering*, edited by G. Chen, pp. 65–85, World Scientific Publishing Co., Hackensack, NJ.

Pan, M., and E. F. Wood (2006), Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter, *J. Hydrometeorol.*, *7*, 534–547.

Reichle, R. H., and R. D. Koster (2002), Land data assimilation with the ensemble Kalman filter: Assessing model error parameters using innovations, in *Developments in Water Science—Computational Methods in Water Resources*, Vol. 47, pp. 1387–1394, Elsevier, New York, NY.

Reichle, R. H., and R. D. Koster (2003), Assessing the impact of horizontal error correlations in background fields on soil moisture estimation, *J. Hydrometeorol.*, *4*(6), 1229–1242.

Reichle, R. H., and R. D. Koster (2004), Bias reduction in short records of satellite soil moisture, *Geophys. Res. Lett.*, *31*, L19501, doi:10.1029/2004GL020938.

Reichle, R. H., D. McLaughlin, and D. Entekhabi (2002a), Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, *130*(1), 103–114.

Reichle, R. H., J. P. Walker, R. D. Koster, and P. R. Houser (2002b), Extended versus ensemble Kalman filtering for land data assimilation, *J. Hydrometeorol.*, *3*(6), 728–740.

Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe (2007), Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth observing system (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR), *J. Geophys. Res.*, *112*, D09108, doi:10.1029/2006JD008033.

Sharif, H. O., W. Crow, N. L. Miller, and E. F. Wood (2007), Multidecadal high-resolution hydrologic modeling of the Arkansas-Red river basin, *J. Hydrometeorol.*, *8*, 1111–1127.

Slater, A., and M. Clark (2006), Snow data assimilation via an ensemble Kalman filter, *J. Hydrometeorol.*, *7*, 478–493.

Van Den Hurk, B. J. J. M., L. Jia, C. Jacobs, M. Menenti, and Z.-L. Li (2002), Assimilation of land surface temperature data from ATSR in an NWP environment- A case study, *Int. J. Remote Sens.*, *23*, 5193–5209.

Zhou, Y., D. McLaughlin, and D. Entekhabi (2006), Assessing the performance of the ensemble Kalman filter for land surface data assimilation, *Mon. Weather Rev.*, *134*, 2128–2142.

W. T. Crow, Hydrology and Remote Sensing Laboratory, USDA/ARS, Beltsville, MD, USA.

C. L. Keppenne and R. H. Reichle, Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Code 610.1, 8800 Greenbelt Road, Greenbelt, MD 20771, USA. (rolf.reichle@nasa.gov)